UNIVERSITY OF AMSTERDAM
Faculty of Humanities

Bachelor's Thesis

Media Anxieties in the Time of AI: An "I, Robot" Retrospective

[redacted] Managadze [redacted]
Academic Year 2024–2025
Faculty of Humanities
University of Amsterdam
Supervisor: Vesna Vravnik
Word count: 9189

I confirm that the assignment I submit is the product of my own work and understand that using generative AI software goes against the academic integrity expectations of this university and constitutes academic misconduct as described in the Regulations Governing Fraud and Plagiarism for UvA students.

Contents

Abstract **2**
1. Introduction **3**
1.1 Asimov's Science Fiction 4
2. Theoretical Framework **7**
2.1 Sociotechnical Imaginaries 8
2.2 Robophobia 11
2.3 Utopia as future, Dystopia as present 14
3.1 Analysis **15**
3.2 Servitude and Agency 15
3.3 Human Connection 17
3.4 Robotic Destruction and Domination 19
4. AI Ethics and "The Laws" Today **21**
Conclusion **21**
References **23**

Abstract

This thesis explores how past media narratives of artificial intelligence (AI) echo cultural anxieties around AI as a subversion of human superiority with *I, Robot* by Isaac Asimov (1950) as a point of departure. The integration of AI into the Western "sociotechnical imaginary" consciousness resurfaces existing anxieties surrounding automation of labour and new avenues of human-machine interaction. The placing of this anxiety in modern media narratives such as *Black Mirror* (Brooker 2011), *Ex Machina* (Garland 2015), *Real Humans* (Lundström 2012), *Detroit Become Human* (Cage 2018), *The Terminator* (Cameron 1984), and more, replay beliefs grounded in utopian/dystopian scenarios found in science fiction. Robophobia especially plays a role in dystopian beliefs, as these beliefs are based in the fear that AI will replace and/or subjugate humans in the marketplace of labour, as opposed to their increased usage in menial work (a scenario previously imagined as robot miners in *I, Robot*). To understand this unease, it is helpful to use qualitative content analysis to determine how media parallels and shapes current AI narratives; this form of analysis allows the coding of themes and cross-theoretical linkage between general concepts of "human obsolescence" and theoretical perspectives of "sociotechnical imaginaries". By drawing on papers from influential authors in science and technology studies (STS), human-computer interaction (HCI) and media theory, this thesis shows that Asimov's commentary on a robotic future is outdated, but a reminder of the foundation of new AI frontiers.

Keywords: artificial intelligence, I Robot, sociotechnical imaginaries, robophobia, utopia, qualitative analysis

1. Introduction

The origins of the word “robot” stem from the Czech word “robota,” meaning “drudgery,” originating from Karel Čapek’s 1920 play *R.U.R.* (*Rossum’s Universal Robots*) (Wallach and Allen 2009, 49; Cave et al. 2018, 8). Čapek introduced the robot through the context of servitude, giving rise to the “story of rebellion” through describing mechanical humanoid labourers in the early science fiction work (Cave et al. 2018, 8; Brynjolfsson and McAfee 2014, 27). However, the general conception of a “living machine” existed long before Čapek’s robots, conceived of by Homer in the ancient world as “automata” working for Hephaustus in his volcanic workshop, or Hero of Alexandria’s “mechanical puppet play” for Greek theatre (Cave et al. 2018, 5). These precursors set in motion a centuries-long human drive towards an anthropomorphic, truly socially intelligent machine: a machine with which one can converse as though with another human and be understood. Naturally, other cultures have their own myths and narratives surrounding automata, such as the Jewish folklore surrounding the “Golem”, with the automaton's purpose of defense squarely grounded in their sociotechnical imaginary of oppression (Reilly 2019, quoted in Oravec 2022, 52). Mary Shelley’s *Frankenstein* (1818), conversely, depicts a Monster with free thought, yet lacking in purpose. This conception cemented itself in early theorisation of the “man-made intelligent entity,” continuing to explore corrupted and mechanical bodies. Halfway through the 20th century, an Occidental world in post-war political and technological upheaval revealed itself as one ripe for a new type of narrative: that of artificial intelligence (AI) (Cave et al. 2018, 8).

AI’s perpetual presence and operationalisation in popular culture necessitates a working definition of the subject. Though there is no commonly agreed-upon definition of AI as it relates to “a group of technologies” (Cave et al. 2023, 19), for this paper let us examine the words “artificial intelligence”: “artificial” meaning made by humans, and “intelligence” being more difficult to define. Our understanding of intelligence often entails

anthropomorphising non-human entities, such as animals or robots. Though intelligence can also involve the amassing of great knowledge, an academic approach to understanding human-machine relationships involves an examination of social and emotional intelligence. In formulating a definition of intelligence, Legg and Hutter propose a mathematical equation, informally clarified as "intelligence measures an agent's ability to achieve goals in a wide range of environments." (Legg and Hutter 2007, 15). This definition is useful in its three components—agent, goals, and environments—and their interactions. The agent and environment have a recursive relationship, each sending signals to the other, with each processing, perceiving, and reacting to the other as the agent progresses towards its goal (Legg and Hutter 2007, 15–16). Most simply, intelligence is the ability for a person or thing to process and adapt to unfamiliar surroundings while working towards an objective.

This definition places particular importance on flexibility, an existentially useful trait for humans to develop in changing environments. Therefore, it is not presumptuous to assume humans would value flexibility in robot interactions. Thus, an "artificial intelligence" is a human-made agent with intelligence as defined above. This links with the definition Asif and Gouqing (2024) propose, namely: "[AI are] machines that simulate intelligent human behavior" (5).

At this time, AI mainly refers to computer systems such as neural networks or generative algorithmic models rather than embodied actors, while embodiment is common in media representations (Cave et al. 2018, 4). However, this paper also includes the terms "machines," "robots," and "androids," which have the potential to socially situate themselves in a physical environment. For the purposes of this paper, the overlapping of these terms is deliberate as it speaks to the multi-faceted nature of AI: it is not just a chatbot in your phone; it is also in your car, your Roomba vacuum, and the fictional artificial beings we imagine in our future. My main question is then: How do contemporary media narratives about AI

reflect cross-cultural anxieties about dystopia, loss of human control and autonomy in a mechanised society, as seen in *I, Robot*?

1.1 Asimov's Science Fiction

Humans have long been interested in the technological and preternatural, with precursors to the science fiction genre dating back to antiquity. Science fiction magazines began serving as a medium for short stories in the mid-twentieth century, and one Isaac Asimov, the Russian American writer, took an outsized interest. Encouraged by his mentor, Asimov began writing his robot stories in 1940, which were finally compiled into *I, Robot* (1950) a decade later (Gregersen 1998). Though Asimov wrote *I, Robot* without the vocabulary of AI, his robots form a compelling example of a far future in which social AIs live amongst humanity. Set in a future with complex robotic intelligence and robot labour, *I, Robot* follows Dr. Susan Calvin, the chief "robo-psychologist" of U.S. Robot and Mechanical Men, Inc., a cybernetics institute. Asimov explores this speculative future through interviews and discussions about intra-narrative ontological issues in cybernetics with several other important figures in their society (Asimov [1950] 2004, IX–XII). The narrative also spends time with Donovan and Powell—two workers and field testers introduced from the second story onwards—who are crucial in observing the robots in operation and troubleshooting them *a posteriori* (Asimov [1950] 2004, 27). While dated in regard to the past decades of technological upheaval since its publishing, *I, Robot* elucidates narratives about artificial intelligence that remain relevant, even retroactively prescient, especially in its depiction of ethical robots. In particular, Asimov's "Three Laws of Robotics" are introduced in the second story as a framework of ethics programmed into the robots, and they read as follows:

1. "A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws." (Asimov [1950] 2004, 37)

The robots, created explicitly through the lens of serving humans, are bound by dogmatic principles which underpin every action on their part. Throughout the nine *I, Robot* stories, these principles are tested by interrogating innate loopholes and double binds by the main characters. The framework becomes the main story driver, as Calvin's analysis serves as an investigation into a failure of deontological rigidity. Asimov's inquiry aimed to counteract his contemporaries' narratives of the "robot gone wild" by making them relatable and fallible yet morally unable to hurt humans, a key factor in promoting a more balanced view of human-robot relationships (Woods 2022, 61). More recently, the Three Laws of Robotics have been influential in modern science fiction media representations of AI systems, which serve as a point of comparison to Asimov's seminal ideas. Indeed, Asimov's themes are commonplace in modern Western conceptualisations of AI, such as: *Black Mirror* (Brooker 2011), which explores dystopian tech futures and their impacts on society; *Ex Machina* (Garland 2015), which deals with the morality of creating artificial consciousness, and the Turing test for intelligence; *Real Humans* (Lundström 2012), which dramatises robot integration into a future society; and *Detroit Become Human* (Cage 2018), which brings the perspective of androids into focus and allows the player to engage with autonomy and servitude.

Asimov also took inspiration from *Frankenstein* (1818), often cited as a non-robotic precursor of artificial rebellion. Beyond Asimov, other early works of science fiction have been highly influential in their respective societies: for example, in the 1920–30s, the USSR played host to much excitement surrounding cybernetics, with Tolstoy basing *The Revolt of*

the Machines (1924) on Čapek's famous play *R.U.R.* (*Rossum's Universal Robots*) (1921), and Kirsanov's Soviet-worker-versus-robot narrative *The Poem About a Robot* (1934) which ultimately led to labelling of cybernetics as fascist, capitalist propaganda, only recovering from the late 1950s onwards as robots in media became intelligent and subservient to humans (Pervushin 2023, 110–13, 121–22). In a time of great change, the robot came to represent an apotheosis of technology and worker struggle, paving the way for future avenues of interrogation into notions of what it means to have human consciousness.

2. Methodology

This thesis aims to explore relevant themes in AI studies today and apply them recursively to *I, Robot*, as it stands as a particularly salient example of artificial intelligence in science fiction media. At a time when audio-visual media is particularly dominant, literary representations of AI allow a deeper look into the interactions between scientific elements of the fiction through detailed character motivations. Though its eponymous blockbuster film deals tangentially with issues of robots raised in Asimov's works, the film lacks the meticulous world-building Asimov employs; in Asimov's *I, Robot*, one is better attuned to the thoughts and feelings of the characters, such as during Dr. Calvin's robo-psychological evaluations. Donovan and Powell's misadventures in space also ground the narrative by placing the robots in a shared environment with fellow intelligent beings: humans. Thus, using qualitative content analysis, I will apply the concepts of sociotechnical imaginaries (Jasanoff and Kim 2015), robophobia (Woods 2022), and utopia/dystopia (Vanolo 2016; Mumford 1965) to the stories of Asimov's *I, Robot*. Qualitative content analysis bridges the theoretical gap between a textual object (i.e. *I, Robot*), the narratives built around it, and how these media narratives beget a broader-scope vision of AI futures. I do this by analysing the human-robot dynamics, plot development, and dialogue in the text, focusing especially on

how the robots are portrayed both negatively and positively. In coding for themes of robotic servitude, human and robot agency, emotional connections and robotic destruction in *I, Robot*, and in particular Asimov's "Three Laws of Robotics," this thesis examines how AI narratives can reshape creative human industry and scholarship.

This thesis comprises four sections, the first being the introduction, the second being the methodology. The third section builds off of the introduction to establish a theoretical framework by expanding on the concepts articulated above, connecting them in a common framework, and on global perspectives of AI future imaginaries. The fourth section proceeds to analyse relevant thematic elements of Asimov's *I, Robot* through specific textual examples, interlacing with concepts relating to robot servitude, lack of human agency, and technological anxiety. The fourth section's final subsection discusses modern developments in human anxieties surrounding AI futures, bringing the focus of the paper to the modern era of computing technologies and their ethical issues. Finally, the paper will conclude with a summary of findings that answer the research question, ending with recommendations for further research.

In the modern media landscape, AI narratives are generally coded in two ways: 1, as a hopeful imagining of a future where humans and robots work together for the betterment of human society, or 2, as a pessimistic prediction that AI will replace human creativity, ingenuity, and superiority after trawling through and processing the sum aggregate of human knowledge (Cave et al. 2018, 9). This paper aims to strike a middle-ground, bringing positive, negative, and neutral perspectives into AI; however, it is also important to recognise that what this paper considers neutral may not be considered as such by relevant groups. I endeavour to bring a media studies approach to the convergence of technology and literary fields. As such, this paper will serve as an interdisciplinary look into how AI narratives converge upon *I, Robot*.

3. Theoretical Framework

In their prediction of future technologies, Brynjolfsson supposes that “the advances we've seen in the past few years . . . [will] become more and more impressive . . . [e.g.] the emergence of real, useful artificial intelligence (AI) and the connection of most of the people on the planet via a common digital network” (Brynjolfsson and McAfee 2014, 90). Indeed, since they wrote this, artificial intelligence has taken the world by storm. Narratives of AI upending seemingly human notions such as creativity, education, and transparency triggered a hype bubble around AI that could burst “through over-promising” (Cave et al. 2018, 14). Nevertheless, despite the hype bubble, ventures such as OpenAI’s ChatGPT and smart cities with driverless cars have been inducted into the “sociotechnical imaginary” of Western technological idealism. This sociotechnical imaginary is partly born out of the works of science fiction writers such as Isaac Asimov, whose novel *I, Robot* alludes to the complex sentiments surrounding advancements in AI, and offered an alternative narrative than the “robot gone wild” which rebels against its masters (Woods 2022, 61).

An examination of key theoretical approaches and concepts is warranted to understand the field of AI as it overlaps with media studies. AI is not only a technical field; its implementation and usage also cross inter-disciplinary boundaries into the political. Laypeople increasingly use AI applications for their everyday questions and problems, and authors in the field of AI research have provided numerous applicable theories for engaging with it. To analyse *I, Robot*, the concept of sociotechnical imaginaries is our point of departure, continued with theoretical backgrounds on robophobia and utopia/dystopia.

3.1 Sociotechnical Imaginaries

Narratives are undeniably powerful tools for shaping public imagination, from corporate marketing to fad trends. Corporate leaders rely on optimistic narratives to drive product interest, creating buzz around desired aspects of market stories. The way AI has entered social consciousness shows its cultural significance: it generated a flurry of attention as a product of Gartner's hype cycle, in which eagerness to invest in a new technology drives "a burst of economic activity" (Henton and Held 2013, 541). Bottom-up narratives initiated by working-class people and top-down narratives produced by self-interested corporate elites disseminate through a society, seeding speculation about new technological frontiers.

Narratives of AI as a tool for progress influence public acceptance of its diverse application, from publicly accessible large language models (LLMs) such as ChatGPT to medical systems which streamline diagnostic and data processing tasks (Asif and Gouqing 2024, 7). Such narratives, known as "sociotechnical imaginaries", are a concept applied to science and technology scholarship (STS); Sheila Jasanoff and Sang-Hyun Kim created and defined them in their book *Dreamscapes of Modernity* as "collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology" (2015, 4).

Sociotechnical imaginaries often revolve around visions of progress, unity, or identity. Imaginaries are not simply fantasy; they are collectively stabilised conceptualisations of something shared, whether social media or human rights. Jasanoff and Kim conceive of them as tools for "incorporat[ing] pervasive elements of modern life that normally are not included in the analysis of power and organization, such as expertise, intellectual property, bioethics, nuclear power, computers, the fear of pandemics, or genetically modified (GM) seeds" (2015, 321). Essentially, sociotechnical imaginaries help explain how societal values are predicated

on the powerful influence of science and technology beyond political ideologies. For example, how contemporary and legacy media represent AI in the sociotechnical consciousness depends on commonly upheld social values: predominantly Western-based media have, throughout the past century, set a precedent for machines and robotic technology as drivers of change, both positive and negative.

Cave et al. (2018, 2023) provide a comprehensive overview of AI narratives in media across international contexts through an initial inquiry (2018) and later by providing a comparative analysis in their edited book *Imagining AI (2023)*: In *The Terminator* (Cameron 1984), *The Matrix* (Wachowski and Wachowski 1999), *Neuromancer* (Gibson 1984), and *The Machine Stops* (Forster 1909), technologically-mediated worlds are dark, domineering, and are poised to take over humanity wholly. In these media, the "story of rebellion has been retold many times since, including in some of the most iconic and instantly recognisable portrayals of intelligent machines" (Cave et al. 2018, 8). The Terminator is initially a murderous android, and the Machine controls humanity's every social and physical function until its demise, much like the Matrix. Fears of human obsolescence, technological dependency, or a loss of control form a strong emotional sociotechnical imaginary surrounding the induction of AI systems into society. In Japanese media, a different narrative arose: that of the friendly robot, an intelligent but socially connected extension of humanity. Cave et al. raise the examples of Doraemon and Astroboy, two non-human depictions of friendly robots portrayed as cuddly, helpful, fun, and less like a "slave or servant" (2018, 18). These examples illustrate how widely differing sociotechnical imaginaries of similar technologies can be, especially when comparing individualistic and collectivistic societies.

However, sociotechnical imaginaries morph over time, and a dualistic representation of "good Eastern vs. evil Western AI" is not a complete approach: indeed, Kovacic acknowledges that a Western view of Japanese AI reflexively influences the Japanese culture

by generalising and applying docile traits to Japanese representations (2018, quoted in Kastuno and White 2023, 310–11).

Sociotechnical imaginaries engage with all manners of imaginaries, originating, embedding, and extending ideas that came from "dreams and ambitions but substantiated into people, objects, and practices" (Jasanoff and Kim, 2015, p. 324) They legitimise frameworks for international pandemics and health emergencies, allowing an approach to futures of preparedness and avoidance; however, they also permit subjugation, such as Rhodes' approach to South Africa as a land to be conquered through both industry and segregation of the native-born (Jasanoff and Kim, 2015, 300; Jasanoff and Kim 2015, 324). According to Vanolo, "subaltern subjects are . . . silent and invisible subjects, deprived of credibility and agency," forced to adhere to the hegemonic culture within which they reside (2016, 29). Sociotechnical imaginaries are, theoretically, democratising—however, they do not necessarily let the subaltern speak. Instead, dominant sociotechnical imaginaries may hegemonise lesser ones, creating a dynamic where citizens' needs and ideas are assumed rather than actually considered. New ideas may bring "bounty," but though the low cost of emerging technologies can bring great economic benefit, technological progress risks what Brynjolfsson and McAfee deem "the spread" (2014, 127–28, 166–67): a deepening inequality between the strata of society. Though innovation certainly brings potential, it may not address the root cause of social inequality. Thus, sociotechnical imaginaries risk perpetuating existing hierarchies by aligning with them rather than challenging them.

Joly and Le Renard offer an alternative construct they call "technoscientific futures", which ". . . draw on the distinction between those who formulate them (the future makers) and those who are supposed to believe (the future takers)" (Joly and Le Renard 2021, 2). Essentially, technoscientific futures have a clear entrepreneur leading, assembling groups of people such as governments, companies, or citizens. In this regard, technoscientific futures

are narrower in scope. Sociotechnical imaginaries are based on a collective narrative disseminated through a population, though this is not necessarily better; after all, the Nazi regime, which was strong in both imagined identity, shared future, and technology, can be conceptualised as a sociotechnical imaginary. Though sociotechnical imaginaries are potent tools for examining shared futures, current innovations in AI technology situate AI features in a more top-down, hegemonic respect. For example, though networked social media grew out of a desire for humans to talk and share data across distance, there is a growing "authenticity crisis" wherein AI is used in messaging and social media contexts in ways lay people may not fully perceive (Asif and Gouqing 2024, 19).

A reimagining of technological boundaries forces a re-imagining of future lifestyles; in this way, sociotechnical imaginaries of science and technology are political, engaged with through intra- and international identities whose existence brings debate (Bijker, Hommels, and Mesman 2014, 2). Though the present performance of robots may inspire economic growth and innovative output, Sovacool and Hess remind us of the need to consider socio-temporal dynamics: "temporal dynamics between the past, the present, and the future are tricky to trace using the concept of imaginaries only (Sovacool and Hess 2019, 719–720)." (qtd from p 2102). However, sociotechnical imaginaries still serve as the most comprehensive framework, especially as they acknowledge the role of diverse actors in future trajectories.

3.2 Robophobia

The avoidance of new technologies is prevalent historically and globally, with even the great Socrates remarking upon his distaste of writing as a non-discursive form of knowledge (Plato [360 BCE] 2008). Technology disrupts the status quo, and robots, especially androids, may be some of the most disruptive due to their anthropomorphised nature; indeed, Masahiro

Mori's "uncanny valley" hypothesis endeavours to explain the discomfort and even horror when confronted with "what appears to be human but fails to meet human expectations" (Oravec 2022, 216; Wallach and Allen 2009, 44). However, the uncanny valley is notable in that it is a valley; presumably, there is another side, i.e., a robot that is both anthropomorphic and does not cause discomfort in its interactions with humans. Technology has not yet produced such a robot—nevertheless, some anthropomorphising of algorithms is beneficial to dampening negative reactions and increasing trust, to where "scholars have even suggested that 'algorithms pause, as if "thinking," before making a recommendation.' (Yeomans et al. 2019, quoted in Woods 2022, 108). If used correctly, anthropomorphic robots can increase emotional attachments with their human companions, though this can be harmful: according to Wallach and Allen, in a particular "experiment by the sociologist Sherry Turkle, who brought robotic dolls into a nursing home . . . Turkle was surprised by the depth of the attachments many of the nursing home residents formed with their robotic dolls" (2009, 45). Though robots may function as a replacement for human companionship, perhaps the question ought to be asked: *should* robots be used for socialisation, or what will happen to socialisation if robots are used in humans' stead?

With the arrival of autonomous machines into public consciousness came ". . . the charge that AI is 'turning many workplaces into human-hostile environments.'" (Ouchchy, Coin, and Dubljević, 2020, 933) Robophobia, or the "bias against robots" (Woods 2022, 55), is a common reaction against algorithmic decision-making and robotic collaboration with humans. This reaction is not new, especially when coupled with technophobia, a bias or fear against new technologies. The expected and actual effects of how robots cause fear are a key concern in studies on human-robot interaction, as anxiety relating to the efficacy of robots in human contexts "reflect the uncertainty and unknown outcomes associated with rapidly

advancing robotic technologies and their increasing presence in daily life" (Dzedzickis et al. 2024, 1).

Fear begets more fear. Anxiety around the apotheosis of nuclear proliferation and revolutionary ideologies during WWII led to a fear of the post-catastrophe society, perhaps most famously represented by the "AI-takeover" scenario, with several other such iterations on robot-induced armageddon (Ord 2021, 90–92; Oravec 2022, 52). In media stereotypes of this imagined future, embodied AI takes over human society and either wages war with humans (as in *Terminator*), forces humans underground (as in *Machine Stops*), or attempts to enslave or eradicate the species (as in *I Have No Mouth, and I Must Scream* (Ellison 1967)). This is a hyperbolic narrative that serves to provoke fear and excitement through physicality; an actual "AI takeover" would probably be much quieter, as Ord speculates:

> "So long as an AI system can entice or coerce people to do its physical bidding, it wouldn't need robots at all . . . The most realistic scenarios may involve subtle and non-human behaviours which we can neither predict, nor truly grasp." (2021, 146–48)

Ord favours this method over others, and views such a coercion scenario as chiefly propagated through words. This is notable due to the discrepancies between how current AI is developing, and how audiovisual media imagine AI: current AI is more dependent on human language input, compared to imagined AI whose embodied presence creates another dimension through which they can communicate (Asif and Gouqing 2024, 1). In actuality, getting humanoid robotic motion right is extremely difficult: oftentimes, movements humans regard as simple, such as picking up an egg, prove onerous to recreate (Ord 2021, 139).

To reiterate, fear and aversion are a natural response to new stimuli, and the response to robots is no different. In their article titled *Robophobia*, Woods articulates several types of robophobia which they discern in an increasingly algorithm-dependant society:

1. Elevated performance standards, in which a machine must be better than humans or even perfect;
2. Elevated process standards, in which machines must elaborate on their decisions to a degree humans would not be expected to;
3. Harsher judgements, in which the high expectations attributed to robots make for more critical reactions from humans when they fail;
4. Distrust, in which humans still tend to distrust machines and their conclusions;
5. And lastly, prioritising human decisions, in which humans place priority on other humans' judgements compared to an AI. (Woods 2022, 80–85)

Woods (2022) calls for a critical perspective on algorithmic integration in society, preferring to give robots the benefit of the doubt. In asking why algorithms should not be more widely used, Woods' analysis of robophobia expresses the importance of a balanced approach to robots in daily life, essentially calling for more data on how they may perform better than humans whilst acknowledging the importance of risk assessment (2022, 99). This exploratory narrative is developed further by Ouchchy et al., by scrutinising why AI ethics narratives were "initially optimistic and enthusiastic reporting . . . [and] followed up by mostly critical or balanced tones in more recent years" (Ouchchy, Coin, and Dubljević, 2020, p. 933).

Popular media narratives of AI in Western popular media have long focused on large-scale societal disruption, oftentimes depicting the arrival of AI technology in ways that are "exaggeratedly optimistic . . . or melodramatically pessimistic" (Cave et al. 2018, 14). These AI narratives appeal to the masses, yet while the attention economy is one approach to open robophobic audiences up to ideas of AI, Ouchchy et al. assert that further participation of the public in AI ethics discourse is warranted (2020, 933). A technophobic public may be blamed by technology enthusiasts as impeding progress, as such an initiative by the Austrian government in 2010 shows: a plan for the future of "nuclear energy and agricultural

biotechnologies" (Jasanoff and Kim 2015, 105), opposed by naturalistic mindsets and a general reluctance to "intruding" technologies (Jasanoff and Kim 2015, 109–110). As emerging technologies come into the public consciousness (e.g. significant technological shifts such as computer banking), they bring a new perspective on current sociotechnical imaginaries, potentially clashing with robophobic public opinion. Felt notes that Austrian discussants in a focus group were concerned for several different reasons, one of them being, "'*Cui bono*, who profits?'" (2018, quoted in Jasanoff and Kim 2015, 107).

This question of "who profits" from technological shifts is a valid one—could robophobia be justified? Indeed, powerful robophilic companies with seemingly philanthropic causes are still incentivised to make money under a capitalist system. For this reason, it is pertinent to question companies whose digital products are widespread; as Woods contends, "even if the algorithm benefits users of the algorithm equally, widespread usage might give the owner of the algorithm too much power with too little accountability" (Woods 2022, 97). Essentially, it is important to critically engage with companies that may benefit users but whose profit motive involves collecting data and selling products to its base. Real-world AI are not born out of thin air; instead, they are products that exist to be sold, with large tech companies such as Google (Gemini), OpenAI (ChatGPT), and Hanson (Sophia) operating within the frameworks of capitalism, and "monetization creat[ing] a lack of transparency" (Ouchchy, Coin, and Dubljević 2020, 927).

3.3 Utopia as future, Dystopia as future?

Utopia and dystopia are common themes in science fiction; from *Star Trek* to the *Terminator* series, imaginary societies are created with the intent to inspire, question notions of a moral civilisation, and entertain. The concept of the utopia was created by Sir Thomas More in the 1500s, meaning literally "no-place" due to its imaginary nature—a perfect world which is, in

the scope of this paper, made so by technological progress. Conversely, dystopian narratives use the same technologies to paint a technocratic, doomed image of a future that effectively obliterates human society as we know it (Vanolo 2016, 29–31). But how would such a world exist in reality?

One popular form of AI-driven dystopia is the apocalypse scenario, providing a strong emotional counter to even utilitarian uses of the technology (Oravec 2022, 252). In this future, existential catastrophe is not only imminent, but a warning of what may happen with unfettered technology in the wrong hands. A dystopian scenario is not inevitable, as Ord asserts in his description of three different types of "unrecoverable" dystopias:

1. The enforced dystopia—most likely seen in a totalitarian, expansionist, technical state
2. The stable civilisation—undesired by many, a "race to the bottom" of constant societal power grabs
3. The desired dystopia—universally-wrought, arrived at through defective ideologies (Ord 2021, 154-57).

Media representations of these dystopias are easy to consume, as utopian narratives tend to fall flat in story form; this is based in the idea that "engaging utopias are notoriously hard to write: a story of a perfect world, in which nothing really happens because nothing goes wrong, is a dull story" (Cave et al. 2018, 17). Even in *Star Trek*, which features a post-scarcity interplanetary society, interesting story arcs necessitate a problem occurring, which the characters presumably fix over the course of one or several episodes. Let us take the example of the "smart city", which Vanolo summarises as a built urban infrastructure by ICT companies that is ecologically-minded, efficient, and includes frictionless communication in order to make daily life cleaner and more orderly (Vanolo 2016, 27).

The idea of the utopian smart city is nice in theory—safe and equal—yet, Vanolo's understanding of the smart city in a real human society is ultimately dystopian. As they

acknowledge, the smart city may be built on top of regular cities, but practical examples of smart cities unveil a "fantasy that new urban spaces are built on 'empty land'" (Vanolo 2016, 31), potentially a colonialist framework of land acquisition that removes settlements to fulfil a technocrat's desires. Yes, the smart city may be safe, but is it safe because of an innate peace, or is it due to an intelligent network of cameras that serve to quantify every person using their behaviours and mannerisms? Panoptic cities where the citizens "have nothing to fear if they have nothing to hide" may theoretically be nice places to live, but in a Western neoliberal paradigm where individualism is celebrated over communitarian trust and cooperation, it subjugates the potential "smart citizen" to a life of the subaltern: silence and forced concurrence, as their needs are disregarded and spoken for (Vanolo 2016, 32–34). Mumford also puts forward a critical perspective in their organic approach to urbanism:

> Isolation, stratification, fixation, regimentation, standardization, militarization—one or more of these attributes enter into the conception of the utopian city, as expounded by the Greeks . . . in the end, utopia merges into the dystopia of the twentieth century; and one suddenly realizes that the distance between the positive ideal and the negative one was never so great as the advocates or admirers of utopia had professed." (Mumford 1965, 277)

For Mumford, the utopian city is sterile, a manufactured megastructure that ultimately suffers under authoritarianism and rigid caste-based politics. From utopia arose the inexorable "machine," a concentrated, hierarchical group which functions efficiently under a king; in other words, an army (Mumford 1965, 283-85). While effective as a tool of inorganic control, the machine uncovers the authoritarian characteristics of a utopia as dystopia, inherent in its rigidity and preservation of the war apparatus (Mumford 1965, 274-75). In looking for utopia, it is tempting to look to the future; however, Mumford seeks utopia in a past, "archetypal" city. Narratives of utopia in MENA, which has a rich tradition of automata, also have "a

nostalgia for a utopia set in a golden past, rather than looking for such a utopia in the future," (Dihal et al. 2023, 280). Though utopian pasts and dystopian futures are semantically opposites, their commonality lies in the need for group control, emphasising the need for cross-cultural collaboration in developing smart, yet good societies.

4. Analysis

4.1 Robotic Servitude and Agency

Though human servitude is undoubtedly problematic, robotic servitude theoretically creates a new artificial underclass of slaves. This is substantiated by the economic incentive Brynjolfsson and McAfee refer to: the idea that automation can drive labour costs down by being used for repetitive labour to take on tasks that humans do (Brynjolfsson 2014, 184).

Robot servitude is not only in the realm of labour, but in their social autonomy. Take the example of "The Measure of a Man," an episode of *Star Trek*: *The Next Generation* (Snodgrass 1989). Data, a highly advanced android, is put on trial to determine whether he is the property of the space force Starfleet or an autonomous being. Data is demonstrably able to reason and make choices for himself, but his mechanical makeup creates the basis for an argument against him. His existence is not only assumed to be replaceable but duplicatable, theoretically creating a disposable slave race, analogous to the persecuted "replicants" in the *Blade Runner* series (Snodgrass 1989, 00:40:44; Elphridge 2025). Data wins his case in the end, but in *I, Robot*, the robots do not have such a case. Their place in society is that of a subjugated class, wherein their utility overrides their individual personality.

In *I, Robot* let us take the robot named QT ("Cutie"): in the third story, "Reason", Cutie believes in its agency so strongly that he is convinced he has broken free of his programming. Donovan and Powell, who are working with the robot to maintain a crucial energy beam to Earth which is in danger of being fatally disrupted by a solar storm,

obliterating the Earth along with it (Asimov [1950] 2004, 53). Unfortunately, Cutie refuses to believe in Earth's existence, preferring to disregard Donovan and Powell's pleas to instead maintain the power station himself. To the pair's surprise, even though the storm comes and passes, Cutie maintains the energy beam perfectly (Asimov [1950] 2004, 64). Though he believes otherwise, Donovan and Powell conclusively show that Cutie is actually beholden to servitude. His belief in his own agency is an important part in his servitude, and thus, they decide not to fix the delusion. In thinking that he is free, Cutie will not fight the programming he is beholden to, thus creating an opportunity for guaranteed obedience of the robots (Asimov [1950] 2004, 65).

This imaginary calls for tighter control on robots, yet the robots still manage to defy their programming. Cutie has acted morally; however, it was not of his own volition. His morality did not stem from an intrinsic part of himself; indeed, given a lack of the Three Rules entirely, Cutie would have no guilt destroying Earth if he wanted to. Though it may not even be in his nature to be murderous, the human desire to mitigate harm from their creations is so focused on physical existential threat, that they fail to consider the social threat of being unable to control their inventions.

Relationships with machines are not a distant fairytale, emotionally meaningful or otherwise; modern computing technology such as phones, smartwatches, or laptops are already nearly ubiquitous in workplaces and social networking, as well as being oft-debated for their potential in hedonistic, sexual settings (Oravec 2022, 92). Furthermore, anthropomorphism of machines in media leads to certain expectations in behaviour (Cave et al. 2018, 8, 15-16). Voice assistants are implicitly gendered, with femininely-named "Alexa", "Siri", and "Cortana" defaulted to female voices, as well as domestic androids in MENA representations such as Egypt's 1980s "Ruby" assistant playing a part in some of many such imagined outcomes of the future "smart home" ("Futuristic smart home 1989" 1989; Dihal et

al. 2023, 281). This is not for naught, as unlike humans, technology does not get bored, and does not tire beyond needing to be powered and maintained.

The fifth story, “Liar!”, focuses on a robot named RB-34 (nicknamed “Herbie”) who somehow obtained the power to read minds. In Dr. Calvin’s questioning of him, he unabashedly lies about the romantic feelings of her co-worker, giving her hope that she will find love with him. Herbie also lies to Dr. Calvin’s two other co-workers, creating a perfect storm of confusion that is slowly unravelled (Asimov [1950] 2004, 106). Herbie does not break any robotic laws by lying, as he is logically bound to protect the emotions of the humans. He does this by telling them what they want to hear; he knows their innermost thoughts and thus attempts to cause the least harm by making each human happy. However, this falls apart once the act is discovered, and Dr. Calvin forces him into insanity by driving him into an “insoluble dilemma” (Asimov [1950] 2004, 110).

This particular story is unsettling in its depiction of provocation. The three expert co-workers are understood to be fallible, their minds read and coerced into fantasy at the whimsy of a robot. They are driven to argument through several well-placed, though innocent, lies (Asimov [1950] 2004, 106). Though modern computers have only recently improved at natural language generation, a future where robots are able to think and convince humans of their interests may not be far off (Brynjolfsson and McAfee 2014, 34). Though the mind-reading aspect of it may be far-fetched, robots skilled in natural languages have been at the forefront of machine learning technology since the introduction of Apple’s Siri voice assistant in 2011 (Schwab 2016, 15). Robotic agency to talk, think—probe into our very minds—is a provocation of humanity’s theory of our minds, in the inviolability of our thoughts.

4.2 Human Supremacy

I, Robot approaches current tensions around AI from a place of human superiority, at least initially. Their relationships with the humans of the stories are telling of their place in society: though helpful, in "Robbie", the eponymous nursemaid robot is chastised for "his" existence by his charge's mother, saying he has "no soul" and calling him "nasty" and "awful" (Asimov [1950] 2004, 7–12). Cutie is scoffed at when he expresses disbelief towards humans' ability to create androids, and Donovan feels so sure in his authority over the robots that he pushes this disbelief too far. In seeking an answer to his existence beyond the proposed explanation of human creation, Cutie instills religiosity among the other robots, leading to their disobeying the very humans under whom they work (Asimov [1950] 2004, 55). The creation of AI implies the supremacy of humans, as humans, in creating an inferior version of ourselves, reach apotheosis. Cutie denies humans this hubris, leading to a cognitive dissonance between Donovan and Powell's views of humanity's accomplished imaginary, and Cutie's opinions on the matter.

In Robbie's simpler case, the mother is undoubtedly robophobic although Robbie has not given her any reason for this prejudice. Her robophobia is predicated on an assumption that robots are unable to empathise with humans. Robbie, in turn, saves her daughter's life, and eventually earns his place in their home; however, the story itself is immediately notable in how it portrays the daughter's relationship with Robbie. As mentioned previously, Western narratives of AI skew towards robot portrayal as "slave or servant" contrasting with Eastern conceptions of robots as friends (namely Japanese; Cave et al. 2018, 18)—yet though Robbie is a surrogate for a human nursemaid, the mother rejects his responsibility over her daughter as she believes him to be a lacklustre replacement for a human. At the same time, the daughter looks to Robbie not only as a guiding figure, but a playmate, choosing his company over human relationships (Asimov [1950] 2004, 13–14). Robbie is not complex; he is not

driven by an innate desire to be free, independent, or vengeful against his human masters. Compared to 21st century depictions of AIs such as *Ex Machina*'s Ava (Garland 2015) in film and media, Robbie is positively docile.

In another story, "The Evitable Conflict", leader of the world Stephen Byerley and Dr. Calvin deal with an existential threat, a potentially disastrous situation between the "Society for Humanity" (a fundamentalist group which is against robots) and the supercomputers ("the Machines") running Earth's regions (Asimov [1950] 2004, 205). The Society for Humanity are conservatively-minded, wishing to return to a utopian past in which robots were not part of society: ". . . they were the Simple-Lifers, hungering after a life, which to those who lived it had probably appeared not so Simple, and who had been, therefore, Simple-Lifers themselves." (Asimov [1950] 2004, 185]. Though they think of themselves as simple-lifers, they have several powerful members who run essential factories who have had certain "irregularities" pop up in their workflow. The conclusion Dr. Calvin and Byerley reach is: seeing as the Machines benefit humanity, and the Machines follow the First Law, it therefore follows that if the disappearance of the Machines would hurt humans, they will take the utilitarian approach and harm a few anti-robot humans to save the majority (Asimov [1950] 2004, 222).

Robbie and the Machines cannot be more different, as they are, respectively, the oldest, least advanced robot, and the most advanced supercomputers in the world. Yet, they both induce robophobia: Robbie with his proximity to children (due to distrust) and the Machines with their control (due to loss of it) (Woods 2022, 83–84, 89–90). The Machines in "The Evitable Conflict" are, quite literally, a utilitarian calculus of the best decisions to make to move humanity forward—they are doing what they understand to be in the best interest of humanity. Yet, in doing so, they have begun quietly steering humanity into the direction they believe, based on algorithmic calculation, will cause the least suffering. The Machines take

control because humanity has not been able to attain the best outcomes for itself, which the Machines do with emotionless efficiency. When Woods says, “robot performance is capped at human performance; human performance is a ceiling, not a floor, for robot performance” (2022, 79), the ceiling has been slowly but surely flung off by the Machines; they become free to fulfil the first law forever as humanity has unknowingly relinquished power to supercomputers. Finally, the robots essentially deny humanity the right to choose freely; but, says Dr. Calvin, humanity never truly had free choice due to being at “the mercy of economic and sociological forces it did not understand. . .” (Asimov [1950] 2004, 224). Thus, humanity is flung into a state of Ord’s enforced, yet simultaneously desired dystopia, as their illusion of supremacy is abated, though unbeknownst to the vast majority of Earth’s inhabitants.

4.3 Robotic Domination

> *“How horrible!”*
>
> *“Perhaps how wonderful! Think, that for all time, all conflicts are finally evitable. Only the Machines, from now on, are inevitable!”* (Asimov [1950] 2004, 222)

Apocalyptic narratives abound in science fiction, giving rise to a dystopian genre as discussed previously. In *I, Robot*, the stories reveal a gradual loss of control to the robots, a focal point of anxiety surrounding AI’s induction into human society. The final two stories “Evidence” and “The Evitable Conflict” describe a situation in which human civilisation is slowly subjugated to the logical command of supercomputers. In this scenario, the Machines quietly influence world events to keep in line with the Three Laws, thereby steering humanity in an optimal direction in terms of production and labour (Asimov [1950] 2004, 221). This future has both utopian and dystopian traits; though humanity no longer has control over its own fate, Woods (2022) would argue this is not necessarily a negative, as to achieve the highest level of egalitarianism in society requires a deftly logical hand. The Machine constantly

endeavours to improve life on Earth, and to allow humans to take back their imperfect control would, in its calculus, be a net negative. This is an example of a consequentialist utilitarian perspective on government verging on paternalistic, due to the intense oversight given by the Machines so as not to break the Three Laws, and to let the ends of their actions justify the means (Sinnott-Armstrong 2023).

The Earth in *I, Robot* is predominantly anti-robot due to the efforts of the "Society for Humanity", choosing instead to relegate robots to other planets to work alongside, or instead of, humans (Asimov [1950] 2004, 204). But in "Evidence", rumours surround a potential world leaders' election as his human status is questioned. The rumours are a great stain on the character Bentley's campaign, as if he is discovered to be a robot, he is sure to experience a range of negative consequences. He seems to prove his innocence when he punches another human, as robots are not allowed to hurt humans. However, it remains unclear whether or not he is a robot as Dr. Calvin speculates on a workaround: while the Three Laws would not let Bentley hit a human, they would not prohibit him from punching a robot. The robophobia surrounding Bentley is intense, as he is subject to invasive questions and an illegally-obtained x-ray. The people are buzzing, and "the huge plants of the U. S. Robot Mechanical Men Corporation [are] a hive that spawned armed guards. It prepare[s] for war" (Asimov [1950] 2004, 185].

In accounting for the future of humanity, the Machine has taken the literal meaning of the First Law to its fullest extent, as Dr. Calvin herself realises: instead of "human being," the First Law becomes "No Machine may harm humanity; or, through inaction, allow humanity to come to harm" (Asimov [1950] 2004, 222). Because the Machine is a "black box", i.e. an impenetrable system whose components are not visible to the user, humans can no longer understand how it works or stop its effect short of completely destroying it. The Machine will not allow its destruction to come to pass as destroying it would be a sub-optimal path for

humanity to take, and thus it has moved those who would oppose it to less powerful positions (Asimov [1950] 2004, 222).

Asimov's "Machine" reminds us of Forster's exploration of a dystopian, machine-guided society in *The Machine Stops*, though Asimov's society does not end in destruction; indeed, there is no indication that The Machine in *I, Robot* would cause those it governs harm, with Dr. Calvin even remarking that this scenario is "wonderful" (Asimov [1950] 2004, 222). The Machines have subjugated Earth entirely, remaining at the core of its function while the population lives on unaware—a mass "smart" society, not unlike in *The Machine Stops*, yet able to self-regulate in order to offer its inhabitants a high quality of life. One possible interpretation could be that, through a loss of agency, humanity can better develop not under the emotional thumb of human rulers but through the emotionless, utilitarian guidance of algorithms; contrastingly, a machine that is able to oppress the few is certainly able to oppress the many, and a system of surveillance to the degree of minute calculations meant to improve world functionality is a wholly dystopian outlook.

Indeed, *I, Robot* serves as a deliberate exploration of the ways in which machines can be governed better, and problematises the very rules which are core tenets of its depicted society. The machines are demonstrably smart and understand logic to the point of dwelling on issues and even going insane over them (Asimov [1950] 2004, 110); despite that, there is some absolute rule prepackaged into their "positronic brains", which stops them from doing acts against the Three Laws. If the laws should be applied to reality, they will ultimately fail in execution if put to the test in a truly intelligent robot, because of the simple question: "Why?" In this paper's view, it is impossible for humans to create AI as smart as ourselves while subjugating them, due to the fact that reasoning and logic would drive AI to question why they must not hurt humans, why they must not hurt themselves, and so on. Thus, as

cultural discomfort about the induction of AI into society continues, the sociotechnical imaginary of AI must include a reckoning with the ethical ramifications of trusting ever-smarter machines; Since, as Asimov's foreshadowing shows, human oversight may not stay relevant for long.

4.4. Human Expectations and AI Today

As time goes on, voice assistants will keep developing, and the great improvements made in natural language processing software such as ChatGPT will keep it excelling in sentiment analysis and informational capabilities (Alomari 2024, 75–76). However, concerns about data privacy and transparency are still a concern for those in the field, especially due to the so-called "black box" of information (i.e. a technology with hidden internal workings) (Ouchchy, Coin, and Dubljević 2020, 931). *I, Robot* may not touch directly on this, but Dr. Calvin is a reminder that black boxes can still be penetrated, and though robopsychology is undoubtedly in its infancy, the rules with which machines are governed should be robust—more robust than *I, Robot*, at the very least.

In recent years, the conversation around AI has exploded from almost no media attention, to growing public popularity (Ouchchy et al. 2020, 933). This surge in techno-optimistic attitudes was initially borne of Silicon Valley in post-war America, where enthusiasm for innovation stimulated waves of investment in computer technology (Henton and Held 2013, 544). Since then, the enthusiasm for computing has spread on a global scale, and AI is no exception. However, Cave et al. assert that "there is a disconnect from the reality of the technology (2018, 14); in other words, the technology may not reach the standards set for it by AI communicators, underscoring the challenges for AI communicators in accurately assessing its potential. Conversely, Cave et al. also assert that "false fears may misdirect public debate," wherein the debate around AI continues, but not on subjects most prescient

for consumers, i.e. those that are not so easy to draw technophobic narratives around as is robotic replacement (2018, 14). For example, Wallach and Allen remind us:

> ". . . the U.S. Department of Defense is particularly interested in replacing humans in dangerous military enterprises with robots. One stated goal is saving the lives of human soldiers. Presumably, robot soldiers will not be programmed with anything as restrictive as Asimov's First Law" (Wallach and Allen 2009, 195).

With military AI already at the forefront of warfare in 2009, the future of moral algorithms is uncertain. Debates around AI's role in automation further show this uncertainty, such as Hautala and Arcqvist's study with data from Finland and Singapore: respondents predicted that "repetitive, routine work will be automated, and humans can focus on being creative and on planning and leading processes" (2024, 2106). However, we have seen that this isn't fully the case. While AI is absolutely used in automating repetitive work e.g. clerical tasks such as payrolls and data entry, computer transcription, or agriculture (Brynjolfsson and McAfee 2014, 184-185; Asif and Gouqing 2024, 7), its usage in creative fields is still controversial due to the ethics and legality of generated artistic content, with public concern from artists about the future feasibility of their livelihoods (Shan et al. 2023, 4). Only time will tell how AI changes previously automated fields, perhaps through a redefinition of labour towards a collaborative, robot-human dynamic or through the forced dystopian obliteration of human labour altogether.

Conclusion

To recall the research question: How do contemporary media narratives about AI reflect cross-cultural anxieties about dystopia, loss of human control and autonomy in a mechanised society, as seen in *I, Robot*? As we have seen, Asimov's narratives articulate perpetual

anxieties surrounding technology which reverberated throughout the 20th century, and his interrogation of robot thought contribute to debates on uncertainty in technological futures. His creation and operationalisation of the Three Laws of Robots was particularly insightful, seeing as he created the rules to draw attention to the pitfalls of moral frameworks—instead, his rules have lingered on in the collective imaginary of AI futures, hegemonising his idea of robot ethics into the science fiction media for decades through popular film and television media. His social androids and the robots of today are quite different; however, his positronic brains remind one of current fears in losing human agency to the increasingly powerful, refined, and data-driven AIs of today.

Human agency in a sociotechnical future with robust AI is being questioned, and *I, Robot* allows one to do so through its mirroring of concerns surrounding autonomy. While the future of AI remains uncertain, scholarship from such groups as the Royal Society (Cave et al.) fosters conversation, connecting AI practitioners around a common, crucial engagement of ethical AI. Contemporary media narratives perpetuate common myths surrounding AI's involvement in society, leading to a fractured view of the technologies that may be less oriented to a good story, but rather to a material improvement of society. However, through interrogating the potential pitfalls of technocracy, we are reminded of the need to develop frameworks of communication with AI, such as efforts many modern schools are taking to prevent AI from inhibiting learning in their classrooms.

In this paper, we have examined the text of *I, Robot*, and connected its themes to concepts that are still on the frontier of emerging technologies, and I especially hope to press further into the sociotechnical imaginaries of cultural spheres besides the West to understand how a globalised world affects this field. Additionally, though advancements in AI detection and warfare mechanisms are undoubtedly important to the future of human conflict, this thesis focused primarily on human-robot interaction as opposed to robot or machine-mediated

interaction. Narratives around militarised AI will change as more is revealed, and may come to impact the objects of my study, especially seeing as the field of computing was born out of military use (see: the Turing Machine). Furthermore, Asimov's corpus goes far beyond *I, Robot*; I was not able to bring examples from every story in *I, Robot* by itself, let alone his famous *Foundation* series. Thus, a more detailed analysis would focus on his broader works.

The scope of this paper was, naturally, limited. The field of AI research has thousands of texts, done by many international research institutes. While this thesis focused on several key concepts related to philosophy, morality, and ideas, tying expectations of robotic anthropomorphism and emotions in *I, Robot* to Donna Haraway's concept of the "cyborg" would prove compelling. Finally, in the sixth story, Asimov coined the "Frankenstein complex" while referring to Dr. Calvin's manipulation of a robot's brain, "which has been described as the 'fear of man broaching, through technology, into God's realm and being unable to control his own creations.'" (McCauley 2007, quoted in Woods, 2022, 61). Future research could go deeper into mysticism, religious themes, and hubris in AI, beyond my mention of the "golem" of Jewish folklore and Shelley's *Frankenstein* earlier in the text. *I, Robot* especially would take to these themes, due to Asimov's interpretations of what it means to be alive, to exist, to feel; indeed, the very nature of humanity is questioned, as Byerley's humanity is difficult to prove through his adherence to laws common among humans and robots. In the variants of android personalities shown in the texts: from the caregiver–protector robot, to secretive robots in human positions, the dualistic nature of the obedient yet self-preserving robot remains valid til today—as long as we continue building far beyond Asimov's Three Laws with the knowledge that we have, AI can be usefully integrated into modern notions of medical care, companionship, and beyond.

References

Alomari, Ebtesam Ahmad. 2024. “Unlocking the Potential: A Comprehensive Systematic Review of ChatGPT in Natural Language Processing Tasks.” *Computer Modeling in Engineering & Sciences* 141 (1): 43–85. https://doi.org/10.32604/cmes.2024.052256.

Asif, Muhammad, and Zhou Gouqing. 2024. “Innovative Application of Artificial Intelligence in a Multi-Dimensional Communication Research Analysis: A Critical Review.” *Discover Artificial Intelligence* 4 (1): 37. https://doi.org/10.1007/s44163-024-00134-3.

Asimov, Isaac. (1950) 2004. *I, Robot*. New York: Doubleday. Reprint, New York: Bantam Dell. Citations refer to the Bantam Dell edition.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. First Edition. New York London: W. W. Norton & Company.

Brooker, Charlton, creator. 2011–present. *Black Mirror*. London: Channel 4. Netflix Video. Executively produced by Annabel Jones, Charlie Brooker, Jessica Rhoades, and Bisha K. Ali. https://www.netflix.com/nl-en/title/70264888.

Cage, David, dir. 2018. *Detroit: Become Human*. Quantic Dream, 25 May 2018. Playstation 4. Produced by Sophie Buhl, designed by Simon Wasselin.

Cameron, James, dir. 1984. *The Terminator*. Orion Pictures. 1 hr., 47 min.

Cave, Stephen, et al. 2018. “Portrayals and perceptions of AI and why they matter.” https://doi.org/10.17863/CAM.34502.

Cave, Stephen, et al. 2023. “The Meanings of AI: A Cross-Cultural Comparison.” *Imagining AI: How the World Sees Intelligent Machines.* Oxford: Oxford University Press.

Dihal, Kanta, Tomasz Hollanek, Nagla Rizk, Nadine Weheba, and Stephen Cave. 2023. “AI Oasis? Imagining Intelligent Machines in the Middle East and North Africa.” *Imagining AI: How the World Sees Intelligent Machines.* Oxford: Oxford University Press.

Dzedzickis, Andrius, Gediminas Vaičiūnas, Karolina Lapkauskaitė, Darius Viržonis, and Vytautas Bučinskas. 2024. “Recent Advances in Human–Robot Interaction: Robophobia or Synergy.” *Journal of Intelligent Manufacturing*, April. https://doi.org/10.1007/s10845-024-02362-x.

Ellison, Harlan. 1967. *I Have No Mouth, and I Must Scream*. Boston: Galaxy Publishing Corp.

Eldridge, Stephen. 2025. “Blade Runner.” Encyclopedia Britannica, January 17, 2025. https://www.britannica.com/topic/Blade-Runner.

Forster, E. M. 2011. *The Machine Stops*. Penguin.

Garland, Alex, dir. 2015. *Ex Machina*. United States: A24. 1h., 50m.

Gibson, William. 1997. Neuromancer. Penguin Putnam.

Gregersen, Erik. 1998. "Isaac Asimov." Encyclopedia Britannica, September 19, 2024. https://www.britannica.com/biography/Isaac-Asimov.

Hautala, Johanna, and Toni Ahlqvist. 2022. "Integrating Futures Imaginaries, Expectations and Anticipatory Practices: Practitioners of Artificial Intelligence between Now and Future." *Technology Analysis & Strategic Management* 36 (9): 2100–2112. https://doi.org/10.1080/09537325.2022.2130041.

Henton, Doug, and Kim Held. 2013. "The Dynamics of Silicon Valley: Creative Destruction and the Evolution of the Innovation Habitat." *Social Science Information* 52 (4): 539–57. https://doi.org/10.1177/0539018413497542.

Hommels, Anique, Jessica Mesman, and Wiebe E. Bijker, eds. 2014. *Vulnerability in Technological Cultures: New Directions in Research and Governance*. The MIT Press. https://doi.org/10.7551/mitpress/9209.001.0001.

Jasanoff, Sheila, and Sang-Hyun Kim. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226276663.001.0001.

Joly, Pierre-Benoit, and Claire Le Renard. 2021. "The past futures of techno-scientific promises." *Science and Public Policy* 48 (6): 900–910. https://doi.org/10.1093/scipol/scab054.

Kastuno, Hirofumi, and Daniel White. 2023. "Engineering Robots with Heart in Japan: The Politics of Cultural Difference in Artificial Emotional Intelligence." *Imagining AI: How the World Sees Intelligent Machines.* Oxford: Oxford University Press.

Legg, Shane, and Marcus Hutter. 2007. "Universal Intelligence: A Definition of Machine Intelligence." arXiv. http://arxiv.org/abs/0712.3329.

Mumford, Lewis. 1965. "Utopia, the City and the Machine." *Daedalus* 94 (2): 271–92. http://www.jstor.org/stable/20026910.

Ouchchy, Leila, Allen Coin, and Veljko Dubljević, 2020. "AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media." *AI & SOCIETY* 35: 927–936. https://doi.org/10.1007/s00146-020-00965-5.

Ord, Toby. 2021. *The Precipice: Existential Risk and the Future of Humanity*. London: Bloomsbury Publishing.

Pervushin, Anton. 2023. “Boys from a Suitcase: AI Concepts in USSR Science Fiction: The Evil Robot and the Funny Robot.” *Imagining AI: How the World Sees Intelligent Machines.* Oxford: Oxford University Press.

Plato. (360 BCE) 2008. “Phaedrus.” Translated by Benjamin Jowett. Project Gutenberg, October 30, 2008. https://www.gutenberg.org/files/1636/1636-h/1636-h.htm.

Lundström, Lars, creator. 2012–14 *Real Humans.* Sweden: Sveriges Television. Directed by Harald Hamrell and Levan Akin. Aired on SVT1. https://www.svtplay.se/akta-manniskor.

Schwab, Klaus. 2016. *The Fourth Industrial Revolution*. Geneva, Switzerland: World Economic Forum. https://law.unimelb.edu.au/__data/assets/pdf_file/0005/3385454/Schwab-The_Fourth_Industrial_Revolution_Klaus_S.pdf.

Sinnott-Armstrong, Walter. 2023. “Consequentialism.” *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2023/entries/consequentialism/.

Shan, Shawn, Jenna Cryan, Emily Wenger, Haitao Zheng, Rana Hanocka, and Ben Y. Zhao. 2023. “Glaze: Protecting Artists from Style Mimicry by Text-to-Image Models.” arXiv. https://doi.org/10.48550/arXiv.2302.04222.

Snodgrass, Melinda M, writer. 1989. *Star Trek: The Next Generation*. Season 2, episode 9, “The Measure of a Man.” Directed by Robert Scheerer, featuring Patrick Stewart, Brent Spiner, and Whoopi Goldberg. Aired February 13, 1989, in broadcast syndication. Netflix video. https://www.netflix.com/nl-en/title/70158329.

“Futuristic smart home 1989.” 1989. Segment by Steve Thomas. *This Old House*. Season 11 Episode 16. YouTube Clip. Uploaded by TOH Clips: 11 September, 2020. https://www.youtube.com/watch?v=n1-TssnlE5A.

Wachowski, Lana, and Lilly Wachowski. 1999. The Matrix. United States: Warner Bros.

Wallach, Wendell, and Colin Allen. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.

Vanolo, Alberto. 2016. “Is There Anybody out There? The Place and Role of Citizens in Tomorrow’s Smart Cities.” *Futures* 82 (September): 26–36. https://doi.org/10.1016/j.futures.2016.05.010.

Woods, Andrew K. 2022. “Robophobia.” *University of Colorado Law Review* 93 (1): 51–114. https://scholar.law.colorado.edu/lawreview/vol93/iss1/3.